Some remarks on value-at-risk optimization

René Henrion*
Weierstrass Institute Berlin, 10117 Berlin, Germany

August 21, 2006

Abstract

We discuss to observations related to value-at-risk optimization. First we consider a portfolio problem under an infinite number of value-at-risk inequality constraints (modeling first order stochastic dominance). The random data are assumed to be normally distributed. Although this problem is necessarily non-convex, an explicit solution can be derived. Secondly, we provide a (negative) result on quantitative stability of the value-at-risk under variation of the random variable. Although reduced Lipschitz properties (in the sense of calmness) may hold true at continuously distributed random variables under suitable conditions, the result shows that no full Lipschitz property (more generally: Hölder property at any rate) can hold in the neighbourhood of an arbitrary continuously distributed random variable. Even worse, this observation holds true with respect to any probability metric weaker than that of total variation.

Keywords: value-at-risk, stochastic dominance,
Hölder continuity, quantile functions

Math Subject Classifications: 90C15

1 Introduction

Despite of a lot of criticism, the value-at-risk continues being a widely used risk measure in financial or engineering applications. We recall its definition here:
Given a one-dimensional random variable Y which is defined on the probability space $(\Omega, \mathcal{A}, P)$, its p-value-at-risk (for $p \in [0,1]$) is denoted by

$$\mathrm{VaR}_p(Y) := \inf \{t \in \mathbb{R} | P(Y \leq t) \geq p\}.$$

*This work was supported by the DFG Research Center MATHEON *Mathematics for key technologies* in Berlin

The shortcomings of VaR are based on modeling as well as on mathematical arguments. The most severe objection from the modeling point of view is that VaR, due to a lack of convexity, does not reward diversification, which is an essential feature in portfolio optimization. As a consequence, VaR fails to be a coherent risk measure (for a critical discussion see, e.g., [18]). Moreover, VaR -constraints used in stochastic optimization do not take into account the degree of constraint violation, which in a financial context could provide valuable additional information. Both these drawbacks seem to be of less importance in non-financial, say engineering applications. For instance, diversification is not a universally useful strategy ('splitting of forces' effect), and in many engineering applications the (monetary) consequence of constraint violation may not be a function of the degree at which such violation takes place. As far as mathematical objections are concerned, the main argument against the use of VaR-constraints is the typical lack of convexity which leads the problems outside the comfortable world of convex optimization. Moreover, VaR does not behave in a stable manner under perturbations of the random variable. Both issues are extensively discussed in [15] when comparing VaR with the conditional value-at-risk.

In the present paper, we want to present two observations concerning these two properties (non-convexity and non-stability). Our aim is not to reason in favour or against the use of VaR (actually both observations point to opposite directions) but to add some more information to this discussion. The first observation shows, that it is possible to obtain an explicit solution to a portfolio optimization problem which is based on an *infinite* family of VaR-constraints (equivalent to a first order stochastic dominance relation). Although the result requires normally (or more general: elliptically) distributed random data, the problem is inevitably non-convex (in contrast to a single VaR-constraint under such distributions). In particular, it is not related to a corresponding formulation based on the conditional value-at-risk. The result and related observations seem to suggest that it might pay to investigate the nature of non-convexity in VaR-Optimization. The second observation is less favourable in that it demonstrates the general non-Hölder continuity of VaR in the neighbourhood of any continuously distributed random variable and with respect to any probability metric weaker than that of total variation. In addition to its importance for VaR-based optimization, this result clarifies some limits for quantitative stability results in optimization problems under probabilistic (or chance) constraints.

2 A continuous family of VaR-constraints under normal distribution

Typically, optimization under VaR-constraints is considered to be difficult due to a lack of convexity. There exist important special cases, however, under which a solution a solution can be determined even in the nonconvex setting. Therefore it might pay to investigate the concrete structure of VaR-constraints before passing to convex risk formulations. To give a first simple example, consider the following portfolio optimization problem, where

$p \in [0,1]$ is a given probability level:

$$\max\{E\langle z, X\rangle \,|\mathrm{VaR}_p(\langle z, X\rangle) \geq L, \ \sum_i z_i = K, \ z \geq 0\}. \tag{1}$$

Here, X is an n-dimensionsal random variable and z denotes an n-dimensionsal decision vector for distributing a capital K on n assets whose random returns are given by the components of X. The objective is to maximize the expected return $E\langle z, X\rangle$ of all assets under the additional risk constraint that the random return $\langle z, X\rangle$ has a p-value-at-risk exceeding a certain level L. In other words, the probability of the random return exceeding the value L is required not to be smaller than p. The following fact is well-known: If X has an n-dimensionsal normal distribution and if $p \geq 0.5$, then the VaR-constraint $\mathrm{VaR}_p(\langle z, X\rangle) \geq L$ defines a convex region in the space of decision vectors $z \in \mathbb{R}^n$ (see [10], [19]). The same result holds true for more general classes of probability distributions like elliptically symmetric distributions (see, e.g., [9]) or log-concave symmetric distributions (see [11]). It can be shown (see [7], Prop. 2.2) that for a certain range of probability levels p strictly smaller than 0.5, the feasibility region becomes necessarily non-convex. However, this is not an arbitrary type of non-convexity. Indeed, by a duality relation shown in [7] (Th. 2.1), the feasible region must then be the complement of some convex set. Along with the given deterministic simplex constraints of the portfolio, this allows to solve (1) by means of so-called reverse convex programming methods (see [8]). This provides an instance, where non-convex VaR-constraints can be successfully treated.

Of course, one might object here, that probability levels $p < 0.5$ are not significant for practical applications, as the safety threshold is typically chosen close to one (where convexity would hold anyway). Therefore, we discuss in the following a slightly more general portfolio model, which is meaningful from the practical viewpoint and where VaR-constraints generate a type of non-convex problem which can still be solved explicitely. More precisely, we study the following problem

$$\max\{E\langle z, X\rangle \,|\mathrm{VaR}_p(\langle z, X\rangle) \geq \mathrm{VaR}_p(Y) \quad \forall p \in [0,1], \sum_i z_i = K, \ z \geq 0\}, \tag{2}$$

where, z, X and K keep the meaning of (1). There are two different features in (2) as compared to (1): first, the deterministic return level L in (1) is replaced now by a 1-dimensionsal random variable Y, which can be interpreted as a benchmark return referring to a previously used portfolio. Secondly, and more importantly, the VaR-constraint is no longer formulated with respect to a single fixed probability level p but rather with respect to all possible such levels. In this way, ambiguity for choosing the right p-value is avoided. Structurally, the difference between (2) and (1) is manifested by the fact that the feasible region in (2) is defined by an infinite family of inequality constraints, thus one deals with a problem of semi-infinite optimization now. Clearly, the feasible region must lack convexity, in general, because already many of the single constraints are non-convex (see previous paragraph).

Before turning to the solution of (2), we digress briefly in order to provide a better motivation of (2) by its relation to the concept of *first order stochastic dominance* (for recent work on optimization problems under stochastic dominance constraints, we refer to, e.g., [1], [2], [3], [4]). Recall that a random variable U stochastically dominates in the first order another random variable V (notation $U \succeq V$), if

$$P(U \leq t) \leq P(V \leq t) \quad \forall t \in \mathbb{R}.$$

Motivated by [2] and [4] (although so-called second order dominance was considered there), we investigate the following portfolio optimization problem:

$$\max\{E\langle z, X\rangle \mid \langle z, X\rangle \succeq Y, \sum\nolimits_i z_i = K, z \geq 0\}. \tag{3}$$

Here, z, X, Y and K have the same meaning as in (2). The first order dominance constraint $\langle z, X\rangle \succeq Y$ is equivalent to requiring that (see, e.g., [4])

$$Eu(\langle z, X\rangle) \geq Eu(Y)$$

for any nondecreasing function u for which these expected values are finite. This means that the expected random return $\langle z, X\rangle$ of the portfolio defined by z is considered to be superior to the expected return of the benchmark Y no matter which (nondecreasing) utility function u the decision maker chooses for measuring the value of return distributions. This property allows to model risk aversion without the need of specifying a particular utility function. As a special case, for the utility function being the identity, one arrives at the conclusion $E\langle z, X\rangle \geq EY$, so the expected return of the chosen portfolio has to exceed that of the benchmark. Note, however, that the dominance constraint is much more restrictive than this simple relation of expected values. A simple reflection on the respective definitions shows that first order stochastic dominance is also equivalent with a continuous family of VaR-constraints:

$$U \succeq V \iff \mathrm{VaR}_p(U) \geq \mathrm{VaR}_p(V) \quad \forall p \in [0,1],$$

so that (3) turns out to be equivalent with (2). This motivates our setting. It is worth mentioning that often second order stochastic dominance constraints are preferred over first order ones due to their preservation of convexity (see [4]). This is coherent with our previous remarks on potential non-convexity of (2). Here, we want to demonstrate for the example of normally distributed data, that the non-convexity of VaR-constraints does not exclude the derivation of an explicit solution.
In the following, we shall use the notation $X \sim \mathcal{N}(\mu, \Sigma)$ to indicate that an n-dimensional random variable X is normally distributed with mean vector μ and covariance matrix Σ. Moreover, we denote the norm induced by Σ as $\|\cdot\|_\Sigma$.

Proposition 2.1 *Let $X \sim \mathcal{N}(\mu, \Sigma)$ and $Y \sim \mathcal{N}(m, \sigma^2)$ be normally distributed random variables of dimension n and 1, respectively. Then, the first order dominance constraint $\langle z, X\rangle \succeq Y$ with respect to some n-dimensional decision vector z is equivalent to the constraints*

$$\|z\|_\Sigma = \sigma \quad \text{and} \quad \langle z, \mu\rangle \geq m.$$

Proof. Passing to scaled random variables, we see that

$$\|z\|_\Sigma^{-1} \langle z, X - \mu \rangle, \ \sigma^{-1}(Y - m) \sim \mathcal{N}(0,1).$$

Therefore, denoting by Φ the one-dimensional standard normal distribution function, we arrive at the equivalences

$$\begin{aligned}
\langle z, X \rangle \succeq Y \iff\; & P(\langle z, X \rangle \le t) \le P(Y \le t) \quad \forall t \in \mathbb{R} \\
\iff\; & P(\|z\|_\Sigma^{-1} \langle z, X - \mu \rangle \\
& \le \|z\|_\Sigma^{-1} (t - \langle z, \mu \rangle)) \le P(\sigma^{-1}(Y - m) \le \sigma^{-1}(t - m)) \quad \forall t \in \mathbb{R} \\
\iff\; & \Phi(\|z\|_\Sigma^{-1} (t - \langle z, \mu \rangle)) \le \Phi(\sigma^{-1}(t - m)) \quad \forall t \in \mathbb{R} \\
\iff\; & \|z\|_\Sigma^{-1} (t - \langle z, \mu \rangle) \le \sigma^{-1}(t - m) \quad \forall t \in \mathbb{R} \\
\iff\; & t(\sigma - \|z\|_\Sigma) \le \sigma \langle z, \mu \rangle - \|z\|_\Sigma m \quad \forall t \in \mathbb{R} \\
\iff\; & \sigma - \|z\|_\Sigma = 0 \quad \text{and} \quad \sigma \langle z, \mu \rangle - \|z\|_\Sigma m \ge 0 \\
\iff\; & \|z\|_\Sigma = \sigma \quad \text{and} \quad \langle z, \mu \rangle \ge m.
\end{aligned}$$

■

Note that the inequality $\langle z, \mu \rangle \ge m$ is just a translation of the inequality $E \langle z, X \rangle \ge EY$ which was mentioned above as a general consequence of the dominance relation $\langle z, X \rangle \succeq Y$.

For the ease of presentation, we consider first the following relaxation of problem (2) where the deterministic portfolio constraints are omitted (we shall come back to the original problem later):

$$\max\{E \langle z, X \rangle \,|\mathrm{VaR}_p(\langle z, X \rangle) \ge \mathrm{VaR}_p(Y) \quad \forall p \in [0,1]\}. \tag{4}$$

Under the normality assumptions of Proposition 2.1 and referring back to the equivalence of (3) and (2), (4) may be reformulated as

$$\max\{\langle z, \mu \rangle \mid \|z\|_\Sigma = \sigma, \, \langle z, \mu \rangle \ge m\}.$$

Now, as the inequality constraint refers to the same expression $\langle z, \mu \rangle$ as the objective function, it can be replaced by an a posteriori check on feasibility. More precisely, it suffices to solve the problem

$$\max\{\langle z, \mu \rangle \mid \|z\|_\Sigma = \sigma\}$$

and to check, whether its optimal solution $\hat{z}$ satisfies the inequality $\langle \hat{z}, \mu \rangle \ge m$ (if so, then $\hat{z}$ is an optimal solution of the previous problem including the inequality constraint, otherwise this previous problem has an empty feasible set).

Clearly, the last problem is nonconvex due to the nonlinear equality constraint $\|z\|_\Sigma = \sigma$. On the other hand, as the objective function is linear, it is equivalent to solve the problem on the convex hull of the constraint set, which evidently amounts to

$$\max\{\langle z, \mu \rangle \mid \|z\|_\Sigma \le \sigma\}$$

and which certainly is a convex optimization problem. A simple calculation based on the first order necessary conditions reveals that (assuming a regular normal distribution for X and, hence, positive definiteness of Σ) its unique solution is given by

$$\hat{z} = \frac{\sigma}{\|\mu\|_{\Sigma^{-1}}} \Sigma^{-1} \mu. \tag{5}$$

The a posteriori feasibility check mentioned above reduces to verifying the relation

$$\sigma \|\mu\|_{\Sigma^{-1}} \geq m,$$

which is easily calculated from the distribution data of X and Y, respectively. We summarize the previous observations as follows:

Theorem 2.2 *Let $X \sim \mathcal{N}(\mu, \Sigma)$ (with $\mu \neq 0$ and Σ regular) and $Y \sim \mathcal{N}(m, \sigma^2)$ be normally distributed random variables of dimension n and 1, respectively. Then, the feasible set of the relaxed problem (4) is empty if $\sigma \|\mu\|_{\Sigma^{-1}} < m$ and, otherwise, (4) has the unique solution (5) realizing the optimal value $\sigma \|\mu\|_{\Sigma^{-1}}$.*

Coming back to the original, non-relaxed portfolio problem (2), things become slightly more complicated from the technical point of view. The reason is that the convex hull argument, used for the solution of the relaxed problem, is not as straightforward. To proceed in the same way as before, one would need the equality

$$\operatorname{conv} \{z \in \mathbb{R}^n | \, \|z\|_\Sigma = \sigma, \, \sum\nolimits_i z_i = K, \; z \geq 0\} = \{z \in \mathbb{R}^n | \, \|z\|_\Sigma \leq \sigma, \, \sum\nolimits_i z_i = K, \; z \geq 0\}. \tag{6}$$

While the inclusion '$\subseteq$' in (6) is obvious, the reverse inclusion does not hold true in general. However, for K large enough, one arrives at the inclusion

$$\{z \in \mathbb{R}^n | \, \|z\|_\Sigma = \sigma, \, \sum\nolimits_i z_i = K\} \subseteq \mathbb{R}^n_+,$$

which can be used to derive (6). Note, that anyway the ratio K/σ has to respect certain lower and upper limits in order to make the feasible set of (2) nonempty. Then, an explicit solution for the complete portfolio problem (2) can be found again from the corresponding Kuhn-Tucker conditions.

Summarizing, we have seen, that for certain special, yet meaningful optimization problems under value-at-risk constraints, the presence of nonconvexity does not exclude a closed form explicit solution. We note that the observations made in this section can be easily generalized from normal distributions to the larger class of elliptically symmetric distributions (analogous to [7]).

3 The Failure of Hölder Continuity for the value-at-risk

An important feature of risk measures is continuity with respect to changes of the underlying random variable. The reason is, that in optimization problems under risk constraints this random variable is not fixed but depends on a decision vector. Moreover, additional approximation effects can arise from discretization. For instance, remedies to the non-convexity of general VaR-constraints could be found for discrete distributions (see, e.g., [12] and [13]) so that it might make sense to replace a possibly continuous original distribution by an empirical approximation. Therefore, it is of interest to analyze the qualitative or quantitative continuity of $\mathrm{VaR}_p(X)$ as a function of the random variable X or its distribution law, respectively. It is obvious that even qualitative continuity of VaR_p cannot be expected without additional assumptions (continuity fails to hold, for instance at X having a discrete distribution). For continuously distributed X, suitable conditions on the probability density or on the growth of the distribution function enable to derive the so-called calmness of $\mathrm{VaR}_p(X)$, which is a Lipschitz-type property of quantitative continuity (see [6], Th. 7 or [14], Prop. 8). To be more precise, for a one-dimensional random variable X defined on the probability space $(\Omega, \mathcal{A}, P)$, we denote its distribution by $\mu_X := P \circ X^{-1}$. Then, calmness of VaR_p at X means the existence of $L, \delta > 0$ such that

$$|\mathrm{VaR}_p(Y) - \mathrm{VaR}_p(X)| \leq L\tilde{\rho}(\mu_Y, \mu_X) \text{ for all } Y \text{ such that } \tilde{\rho}(\mu_Y, \mu_X) < \delta, \tag{7}$$

where $\tilde{\rho}$ denotes the Kolmogorov metric between the distributions of X and Y which is defined by

$$\tilde{\rho}(\mu_Y, \mu_X) = \sup_{t\in\mathbb{R}}\{|\mu_Y((-\infty, t]) - \mu_X((-\infty, t])|\}.$$

Although the estimate (7) looks like a Lipschitz property, it is significantly weaker indeed due to the fact that one random variable (namely X) is held fixed whereas only the second variable Y is allowed to move in a neighbourhood of X. In contrast, a true Lipschitz estimate (which holds, for instance, for the conditional value-at-risk, see [16], Ex. 16) would allow two variables to move independently in a certain neighbourhood. While calmness is already quite a useful property in many respects of optimization problems (e.g., error estimates, constraint qualification for necessary optimality conditions etc.), a true Lipschitz estimate would provide even more useful information, for instance for convergence of algorithms. Unfortunately, as it will be shown below, true Lipschitz estimates cannot be derived for the value-at-risk at any (!) continuously distributed random variable. Even worse, the same negative result holds true for the weaker concept of Hölder continuity and it also holds true not just for the Kolmogorov metric used in (7) but for any probability metric weaker than that of total variation.

To start with some definitions, let ρ be any probability metric on $\mathbb{R}$. We call the p-value-at-risk locally Hölder continuous of rate $\kappa > 0$ with respect to ρ at a fixed one-dimensional random variable X, if there exist $L, \delta > 0$ such that

$$|\mathrm{VaR}_p(Y) - \mathrm{VaR}_p(Z)| \leq L\rho^\kappa(\mu_Y, \mu_Z)$$

holds true for all one-dimensional random variables Y, Z with

$$\rho(\mu_Y, \mu_X) < \delta \quad \text{and} \quad \rho(\mu_Z, \mu_X) < \delta.$$

For any probability measures ν_1, ν_2 on $\mathbb{R}$, let

$$\rho_{TV}(\nu_1, \nu_2) := \sup_{B \in \mathcal{B}} |\nu_1(B) - \nu_2(B)|$$

be the metric of total variation between ν_1 and ν_2, where $\mathcal{B}$ refers to the collection of all Borel subsets of $\mathbb{R}$. If ρ_1, ρ_2 are probability metrics on $\mathbb{R}$, then we call ρ_1 weaker than ρ_2 if $\rho_1 \leq c\rho_2$ for some $c > 0$.

Theorem 3.1 *Let $p \in (0,1)$ and X be an arbitrary one-dimensional random variable having a continuous distribution. Then, the p-value-at-risk is not locally Hölder continuous of any rate at X with respect to any probability metric which is weaker than the metric of total variation.*

Proof. Let $\kappa > 0$ be a supposed Hölder rate of the p-value-at-risk. By assumption, the distribution function

$$F_X(t) := P(X \leq t)$$

is continuous. Define $\bar{x} := \mathrm{VaR}_p(X)$ (note that $\bar{x} \in \mathbb{R}$ due to our assumption $p \in (0,1)$). The continuity of F_X implies that $F_X(\bar{x}) = p$. We consider a sequence of random variables Y_n having distribution functions

$$F_{Y_n}(t) := \begin{cases} F_X(t) & t \in (-\infty, \bar{x}) \\ p & t \in [\bar{x}, \bar{x} + n^{-1}) \\ p + n(t - \bar{x} - n^{-1})(F_X(\bar{x} + 2n^{-1}) - p) & t \in [\bar{x} + n^{-1}, \bar{x} + 2n^{-1}) \\ F_X(t) & t \in [\bar{x} + 2n^{-1}, \infty) \end{cases} \qquad (n \in \mathbb{N}).$$

One easily verifies that the F_{Y_n} are continuous functions which differ from F_X just on the interval $(\bar{x}, \bar{x} + 2n^{-1})$, where they are nondecreasing by definition and due to

$$F_X(\bar{x} + 2n^{-1}) \geq F_X(\bar{x}) = p.$$

Consequently, the F_{Y_n} are continuous distribution functions too. Now, we estimate the distance of total variation between μ_X and μ_{Y_n}. Given any Borel subset B of $\mathbb{R}$, we may decompose it into 4 disjoint parts as

$$B = B_1 \cup B_2 \cup B_3 \cup B_4,$$

where the indices refer to the intersection of B with the four disjoint intervals used in the construction of F_{Y_n}. From the very definitions of F_X and F_{Y_n}, it follows that

$$\mu_X(B_1) = \mu_{Y_n}(B_1), \quad \mu_X(B_4) = \mu_{Y_n}(B_4) \quad (n \in \mathbb{N}),$$

whence

$$\begin{aligned} |\mu_X(B) - \mu_{Y_n}(B)| &= |\mu_X(B_2 \cup B_3) - \mu_{Y_n}(B_2 \cup B_3)| \leq \mu_X(B_2 \cup B_3) + \mu_{Y_n}(B_2 \cup B_3) \\ &\leq 2\left(F_X(\bar{x} + 2n^{-1}) - p\right). \end{aligned}$$

Since B was arbitrary, it follows from the continuity of F_X that

$$d_{TV}(\mu_X, \mu_{Y_n}) \leq 2\left(F_X(\bar{x} + 2n^{-1}) - p\right) \to_n 0. \tag{8}$$

Moreover, we have that

$$\mathrm{VaR}_p(Y_n) = \bar{x}. \tag{9}$$

Indeed, $\mathrm{VaR}_p(Y_n) \leq \bar{x}$ due to $F_{Y_n}(\bar{x}) = p$. On the other hand, the relation $\mathrm{VaR}_p(Y_n) < \bar{x}$ would yield the existence of some $x < \bar{x}$ such that $p \leq F_{Y_n}(x) = F_X(x)$ by construction of Y_n. This, however, results in the contradiction $\mathrm{VaR}_p(X) \leq x < \bar{x}$.

Next, we consider a second sequence of random variables Z_n. First, we put $x_n := \mathrm{VaR}_{p-n^{-2/\kappa}}(X)$ for $n \in \mathbb{N}$ (note that $x_n \in \mathbb{R}$ for n large enough due to our assumption $p \in (0,1)$). The continuity of F_X implies that $F_X(x_n) = p - n^{-2/\kappa}$. Clearly, $x_n < \bar{x}$ and, in particular, $F_{Y_n}(x_n) = p - n^{-2/\kappa}$ for all $n \in \mathbb{N}$. The Z_n are assumed to have distribution functions

$$F_{Z_n}(t) := \begin{cases} F_{Y_n}(t) & t \in (-\infty, x_n) \\ p - n^{-2/\kappa} + n^{-2/\kappa}(t - x_n)\left(\bar{x} + n^{-1} - x_n\right)^{-1} & t \in [x_n, \bar{x} + n^{-1}) \\ F_{Y_n}(t) & t \in [\bar{x} + n^{-1}, \infty) \end{cases} \quad (n \in \mathbb{N}).$$

The F_{Z_n} are continuous functions which differ from F_{Y_n} just on the interval $(x_n, \bar{x} + n^{-1})$, where they are nondecreasing. Consequently, the F_{Z_n} are continuous distribution functions as well. To estimate the distance of total variation between μ_{Y_n} and μ_{Z_n}, consider an arbitrary Borel subset B of $\mathbb{R}$, wich we decompose into 3 disjoint parts as

$$\tilde{B} = \tilde{B}_1 \cup \tilde{B}_2 \cup \tilde{B}_3,$$

where the indices refer to the intersection of $\tilde{B}$ with the three disjoint intervals used in the construction of F_{Z_n}. Similar to the situation above, we have that

$$\mu_{Y_n}\left(\tilde{B}_1\right) = \mu_{Z_n}\left(\tilde{B}_1\right), \quad \mu_{Y_n}\left(\tilde{B}_3\right) = \mu_{Z_n}\left(\tilde{B}_3\right) \quad (n \in \mathbb{N}),$$

whence, by the definitions of Y_n and Z_n,

$$\begin{aligned}\left|\mu_{Y_n}\left(\tilde{B}\right) = \mu_{Z_n}\left(\tilde{B}\right)\right| &= \left|\mu_{Y_n}(\tilde{B}_2) - \mu_{Z_n}(\tilde{B}_2)\right| \leq \mu_{Y_n}(\tilde{B}_2) + \mu_{Z_n}(\tilde{B}_2) \\ &\leq F_{Y_n}(\bar{x}+n^{-1}) - F_{Y_n}(x_n) + F_{Z_n}(\bar{x}+n^{-1}) - F_{Z_n}(x_n) \\ &\leq p - (p - n^{-2/\kappa}) + p - (p - n^{-2/\kappa}) = 2n^{-2/\kappa}.\end{aligned}$$

Since $\tilde{B}$ was arbitrary, it follows that

$$\rho_{TV}\left(\mu_{Y_n}, \mu_{Z_n}\right) \leq 2n^{-2/\kappa} \to_n 0. \tag{10}$$

As a consequence, (8) and (10) may be combined to

$$\rho_{TV}\left(\mu_X, \mu_{Z_n}\right) \to_n 0. \tag{11}$$

Moreover,

$$\mathrm{VaR}_p(Z_n) = \bar{x} + n^{-1}. \tag{12}$$

Indeed, $\mathrm{VaR}_p(Z_n) \leq \bar{x}+n^{-1}$ due to $F_{Z_n}(\bar{x}+n^{-1}) = F_{Y_n}(\bar{x}+n^{-1}) = p$. On the other hand, $F_{Z_n}(t) < F_{Z_n}(\bar{x}+n^{-1}) = p$ for $t < \bar{x}+n^{-1}$ (see definition of Z_n and recall that $x_n < \bar{x}$), whence $\mathrm{VaR}_p(Z_n) \geq \bar{x}+n^{-1}$. Summarizing, the relations (9), (12) and (10) amount to

$$|\mathrm{VaR}_p(Y_n) - \mathrm{VaR}_p(Z_n)| = n^{-1} \geq 2^{-\kappa} n \rho_{TV}^{\kappa}\left(\mu_{Y_n}, \mu_{Z_n}\right) \quad (n \in \mathbb{N}).$$

Now, if ρ is a probability metric on $\mathbb{R}$ weaker than ρ_{TV}, then $\rho \leq c\rho_{TV}$ for some $c > 0$, and the last relation may be extended to

$$|\mathrm{VaR}_p(Y_n) - \mathrm{VaR}_p(Z_n)| \geq n 2^{-\kappa} c^{-\kappa} \rho^{\kappa}\left(\mu_{Y_n}, \mu_{Z_n}\right) \quad (n \in \mathbb{N}).$$

Similarly, (8) and (11) imply that

$$\rho\left(\mu_X, \mu_{Y_n}\right), \rho\left(\mu_X, \mu_{Z_n}\right) \to_n 0.$$

This, however, contradicts the local Hölder continuity of rate $\kappa > 0$ of the p-value-at-risk with respect to ρ.

■

We note that many important probability metrics on $\mathbb{R}$ are weaker than the distance of total variation. Examples are: the total variation itself, the Komogorov metric, more generally, all kinds of discrepancy metrics, the Prokhorov metric, the Wasserstein metric and the Levy metric. By the theorem above, no Hölder continuity of the p-value-at-risk can be expected for all these metrics, at least not in the neighborhood of a continuously distributed random variable.

Finally, we want to recall that, by definition, VaR is the solution 'set' of a very simple optimization problem under a so-called probabilistic (or chance) cosntraint. Therefore, the negative result of Theorem 3.1 has also a bearing on quantitative stability of solution sets to such problems under perturbation of the underlying probability measure. This issue has been intensively analyzed in [5]. The results stated there are limited to reduced Hölder and Lipschitz properties of solution sets, where one parameter (the original distribution) is held fixed and only the second parameter is allowed to move. Theorem 3.1 confirms that no improvement of these results can be expected.

References

[1] D. Dentcheva and A. Ruszczynski, Optimization with stochastic dominance constraints, *SIAM J. Optim.*, 2003, **14**: 548-566.

[2] D. Dentcheva and A. Ruszczynski, Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints, *Math. Program.*, 2004, **99**: 329-350.

[3] D. Dentcheva and A. Ruszczynski, Semi-Infinite Probabilistic Optimization: First Order Stochastic Dominance Constraints, *Optimization*, 2004, **53**: 583–601.

[4] D. Dentcheva and A. Ruszczynski, Portfolio Optimization with Stochastic Dominance Constraints, *J. Banking and Finance*, 2006, **30**(2): 433–451.

[5] R. Henrion and W. Römisch, Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints, *Math. Program.*, 2004, **100**: 589-611.

[6] R. Henrion and J. Outrata, Calmness of constraint systems with applications, *Math. Program.*, 2005, **104**: 437-464.

[7] R. Henrion, Structural Properties of Linear Probabilistic Constraints, *Stochastic Programming EPrint Series* (SPEPS), 2005-13, to appear in: *Optimization.*

[8] R. Horst and H. Tuy, *Global Optimization*, Springer, Berlin, 1996.

[9] Y. Kan, Application of the quantile optimization to bond portfolio selection. **in:** *Stochastic optimization techniques* (K. Marti, ed.), (*Lect. Notes Econ. Math. Syst.*, **513**), Springer, Berlin, 2002, 285–308.

[10] S. Kataoka, A Stochastic Programming Model, *Econometrica*, 1963, **31**: 181-196.

[11] C.M. Lagoa, X. Li, X., M. Sznaier, Probabilistically Constrained Linear Programs and Risk-Adjusted Controller Design, *SIAM J. Optim.*, 2005, **15**: 938-951.

[12] J.S. Pang and S. Leyffer, On the Global Minimization of the value-at-risk, *Optim. Methods Softw.*, 2004, **19**: 611-631.

[13] G. Pflug, Some Remarks on the value-at-risk and the Conditional value-at-risk. **in:** *Probabilistic Constrained Optimization* (S.P. Uryasev, ed.), Kluwer, Dordrecht, 2000, 272-281.

[14] G. Pflug, Stochastic Optimization and Statistical Inference, in [17], Chapter 7, 427-482.

[15] R.T. Rockafellar and S. Uryasev, Conditional value-at-risk for general loss distributions, *J. Banking and Finance*, 2002, **26**: 1443–1471.

[16] W. Römisch, Stability of Stochastic Programming Problems, in [17], Chapter 8, 483-554.

[17] A. Ruszczynski and A. Shapiro (eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, **Vol. 10**, Elsevier, Amsterdam, 2003.

[18] G. Szegö, Measures of risk, *J. Banking and Finance*, 2002, **26**: 1253–1272.

[19] C. van de Panne, W. Popp, Minimum-Cost Cattle Feed under Probabilistic Protein Constraints, *Manag. Sci.*, 1963, **9**: 405-430.